Distribuţia mediilor şi a diferenţelor între medii

Fie o populaţie statistică *N* (N foarte mare), pe care o considerăm ca având o distribuţie normală.

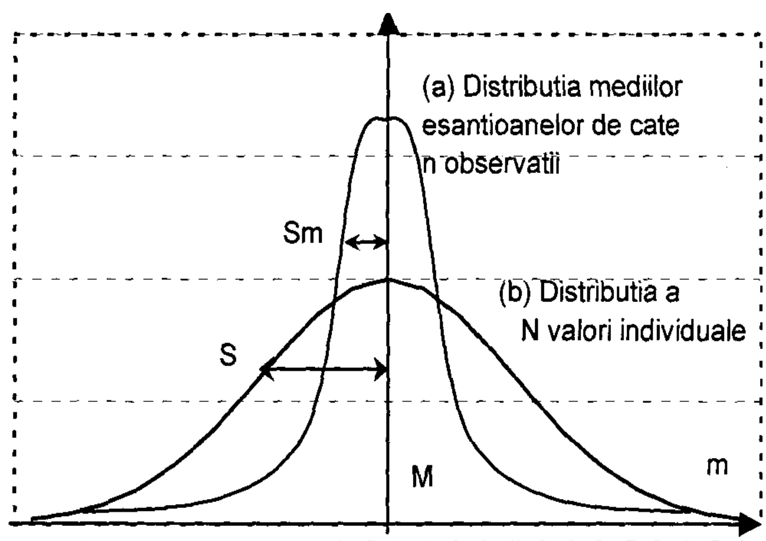
Vom extrage un eşantion de efectiv *n*.

Fie m_1, m_2, m_3 ... mediile găsite pentru diverse eşantioane.

Se studiază fluctuaţia statistică a mediilor eşantioanelor extrase între ele, şi egal repartizate faţă de media M a populaţiei de origine. Se constată că mediile sunt mai puţin dispersate faţă de M, media globală a populaţiei, decât valorile individuale din populaţie.

Distribuţia nou-obţinută în acest mod se numeşte distribuţia mediilor.

Abaterea tip a acestei distribuţii de medii se numeşte abaterea standard a mediei, şi se notează S_m.



Distribuţia mediilor în jurul mediei globale a populaţiei, în comparaţie cu distribuţia valorilor individuale

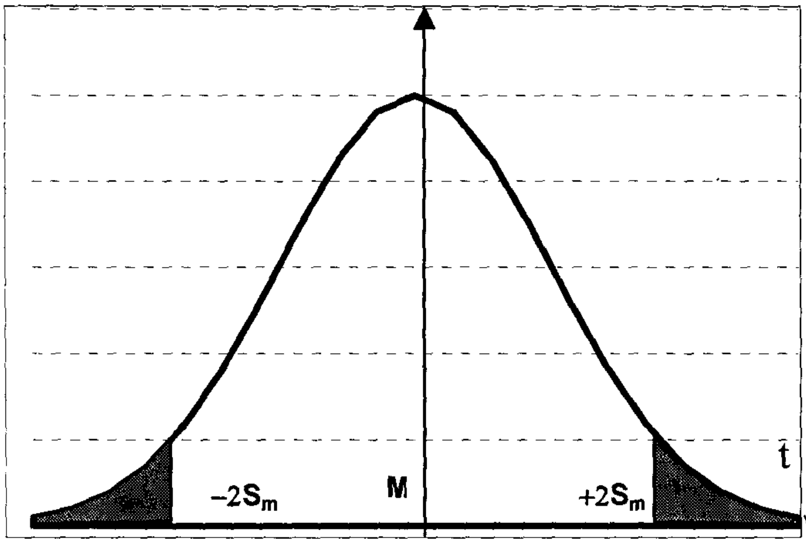
Distribuţia mediilor fiind mai puţin dispersată, abaterea tip S_m este totdeauna mai mică decât abaterea tip S a populaţiei de origine; între cele două mărimi există relaţia:

$$S_m = \frac{S}{\sqrt{n}}$$

Mulţimea mediilor care se pot găsi pentru diverse eşantioane având acelaşi număr de observaţii, extrase la întâmplare dintr-o populaţie de medie M şi abatere standard S, formează aşadar o distribuţie gaussiană de valoare medie M, şi având abaterea tip S_m.

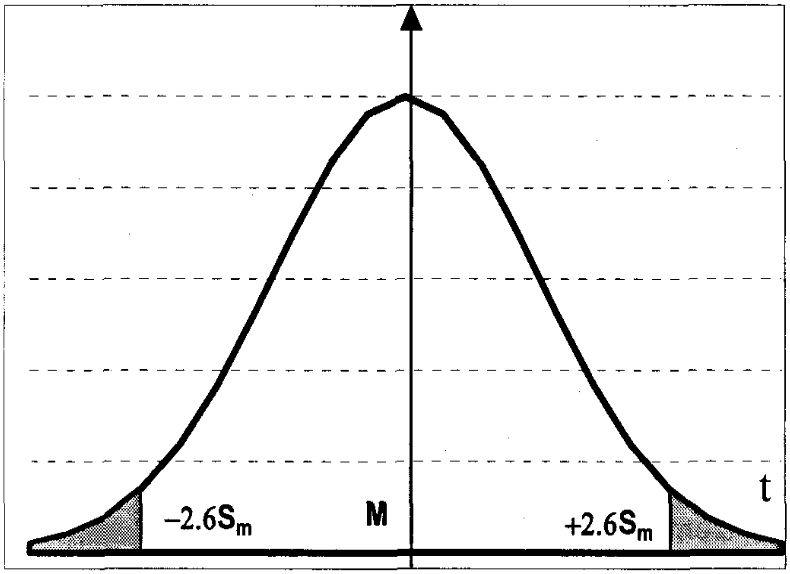
INTERVALUL DE ÎNCREDERE AL MEDIEI

Intervalul corespunzător distribuţiei mediilor, ($M - 2S_m$, $M + 2S_m$), cuprinzând 95,5% din valorile pe care le poate lua media m a eşantionului din mulţimea fluctuaţiilor întâmplătoare, se numeşte interval de confidenţă al mediei cu un coeficient de securitate de 95,5%.



Intervalul de confidenţă al mediei cu un coeficient de securitate de 95.5%

Analog se defineşte intervalul de confidenţă al mediei cu un coeficient de securitate de 99% (Figura 8.47), ca fiind intervalul ($M - 2.6 \cdot S_m$, $M + 2.6 \cdot S_m$) - ne spune că avem 99 şanse din 100 ca media unui eşantion ales să cadă în acel interval.



Intervalul de confidenţă al mediei cu un coeficient de securitate de 99%

DETERMINAREA INTERVALULUI DE CONFIDENŢĂ AL MEDIEI

Dorim să studiem la un eşantion intervalul de încredere al mediei observate, m_0. Nu cunoaştem nici media M, nici S_m , dar presupunem că ştim abaterea tip S a populaţiei de origine.

Câteodată, experienţa ne arată că în practică, oricât de mic ar fi eşantionul, dar suficient de important, distribuţiile de eşantionaj sunt distribuţii sensibil normale. În aceste condiţii, valoarea m_0 găsită pentru m reprezintă valoarea a cărei probabilitate este cea mai mare. În consecinţă, este logic să considerăm că cea mai bună estimare pe care o luăm va fi media M, şi să o substituim în intervalul de confidenţă.

De altfel, abaterea σ a eşantionului reprezintă o estimare a abaterii tip S a populaţiei de origine şi se consideră substituţia lui S cu S_m rezultat din calcul. Abaterea σ a eşantionului va fi o estimare puţin mai mică decât S. Pentru a estima corect S trebuie să luăm o valoare puţin mai mare decât σ al eşantionului. Calculul arată efectiv că cea mai bună estimare a lui S, pe care o vom nota cu S_σ este puţin mai mare decât σ, fiind definită de formula:

$$S_{\sigma} = \sigma \sqrt{\frac{n}{n-1}}$$

Se poate deci utiliza această valoare pentru a calcula S_m , care va fi:

$$S_m = \frac{S_{\sigma}}{\sqrt{n}} = \frac{1}{\sqrt{n}} \cdot \sigma \sqrt{\frac{n}{n-1}} = \frac{\sigma}{\sqrt{n-1}}$$

=> $S_m = \dfrac{\sigma}{n-1}$

Plecând de la valorile estimate ale lui M şi S_m, se va putea exprima intervalul de confidenţă al mediei, care va fi în final:

- $m_0 \pm 2S_m$, cu un coeficient de securitate de 95%;
- $m_0 \pm 2.6S_m$, cu un coeficient de securitate de 99%.

cu $S_m = \dfrac{\sigma}{\sqrt{n-1}}$

<u>Exemplu</u>: Se dozează corticoizii urinari într-un grup de 253 femei cu greutate normală. Se găseşte media m = 4,50 mg/24h şi abaterea tip σ=1,50. Să se găsească intervalul de încredere. Avem:

$$S_m = \frac{\sigma}{\sqrt{n-1}} = \frac{1.5}{\sqrt{252}} = 0.1$$

Intervalul de încredere al mediei este deci:

$m_0 \pm 2S_m = 4.50 \pm 2 \cdot 0.1 = 4.50 \pm 0.2$

=> (4.30 , 4.70) cu un coeficient de securitate de 95%;

$m_0 \pm 2.6S_m = 4.50 + 2.6 \cdot 0.1 = 4.50 \pm 0.26$

⇨ (4.24 , 4.76) cu un coeficient de securitate de 99%.

BIBLIOGRAFIE

1. GEORGESCU GABRIELA, DASCĂLU CRISTINA – “Informatică aplicată şi biostatistică”, Edit. Stef, 2003;
2. JACOBS A.D. –“Medical Biostatics”, Edit. Bucur-Mond, Bucureşti, 1997;
3. MOCANU M.N. – “Curs de informatică medicală”, Braşov, 1996.